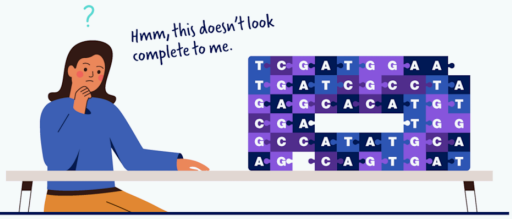
Why was it so difficult to fully complete the human genome sequence?

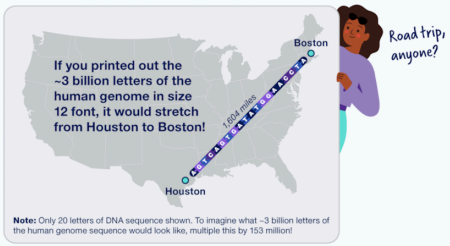
The Human Genome Project ended in 2003, but genomic researchers had not yet determined every last base (or letter) of the human genome sequence. Instead, they had only completed about 92% of the sequence at that time. Why did they stop there?



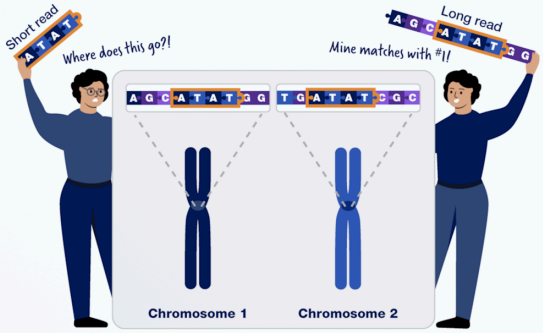
Reason 1

The human genome contains a massive amount of DNA.

The human genome consists of about 3 billion bases in a precise order, each of which can be represented by a letter (G, A, T or C). A genome's sequence cannot be read out end-to-end. Rather, researchers must first determine the sequence of random pieces of DNA and then use those smaller sequences to put the whole genome sequence back together like a massive puzzle.



Reason 2



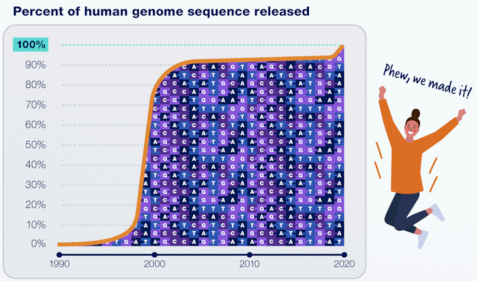
Some parts of our DNA are painfully repetitive.

Some sections of the human genome sequence consist of long, repetitive stretches of letters that are difficult to put in the right place. Over the past two decades, researchers developed new technologies to read longer stretches of DNA — from only about 500 to now over 100,000 letters at a time — which allowed them to assemble the full length of the most difficult repeats.

Reason 3

The first 92% was hard. The last 8% was excruciating.

Those DNA repeats and other obstacles stood between the genomic researchers and the final 8% of the human genome sequence until new laboratory and computational technologies were developed. It took almost twice as long to finish the last 8% of the human genome as it did the first 92%!



Reason 4

The last 8% needed a generation of dedicated genomic researchers with a vision.

Even with new technologies, genome sequencing is still tough, time-consuming work that requires a lot of skill and dedication. The current generation of genomic researchers are true perfectionists and brought everything together to finally complete the human genome sequence.